RDA Finland Meeting & Workshop: Sensitive Research Data Management

Tiedekulma / Think Corner

25.4.2019

RDA EU 4.0 and the National Nodes

RDA EU 4.0 builds on existing networks to create a **European network of nodes** to enhance Member State cooperation and support of RDA in many European countries & vice versa the representation of EU countries in RDA Global


RDA
RESEARCH DATA ALLIANCE
Why attend the next RDA Plenary in Finland?
RDA 14th Plenary
Helsinki | Espoo
Finland
23-25 October 2019
Aalto University
FEDERATION OF FINNISH LEARNED SOCIETIES
FMI
CSC
for Brilliant Minds
UNIVERSITY OF HELSINKI
RDA
RESEARCH DATA ALLIANCE
EUROPE

RDA Finland Meeting

9.30-10.00 Coffee

10.00-10.30 DMPTuuli guidance on managing sensitive data, Information Specialist **Mari Elisa Kuusniemi**, University of Helsinki & Data Management Specialist **Turkka Näppilä**, Tampere University

10.30-10.45 Greetings and summary from 13th Plenary Meeting in Philadelphia, Project Manager **Ville Tenhunen**

10.45-11.00 FAIR guiding principles from the point of view of sensitive data, RDA Finland Coordinator **Heidi Laine**, CSC

11.00-11.30 Sensitive Data in GDPR and National Data Protection Act, Senior Legal Counsel **Maria Rehbinder**, Aalto University

11.30-12.00 Infrastructure for research with sensitive data, Program Director **Antti Pursula**, CSC

12.00-12.30 Concluding discussion

12.30-13.00 Lunch Break (light lunch served for registered participants)

RDA Finland Workshop

13.00-13.30 Introduction to RDA outputs, working groups and interest groups related to sensitive data management, RDA Finland Node Coordinator **Heidi Laine**, CSC

13.30-14.30 Workshop based on the RDA sensitive data related work

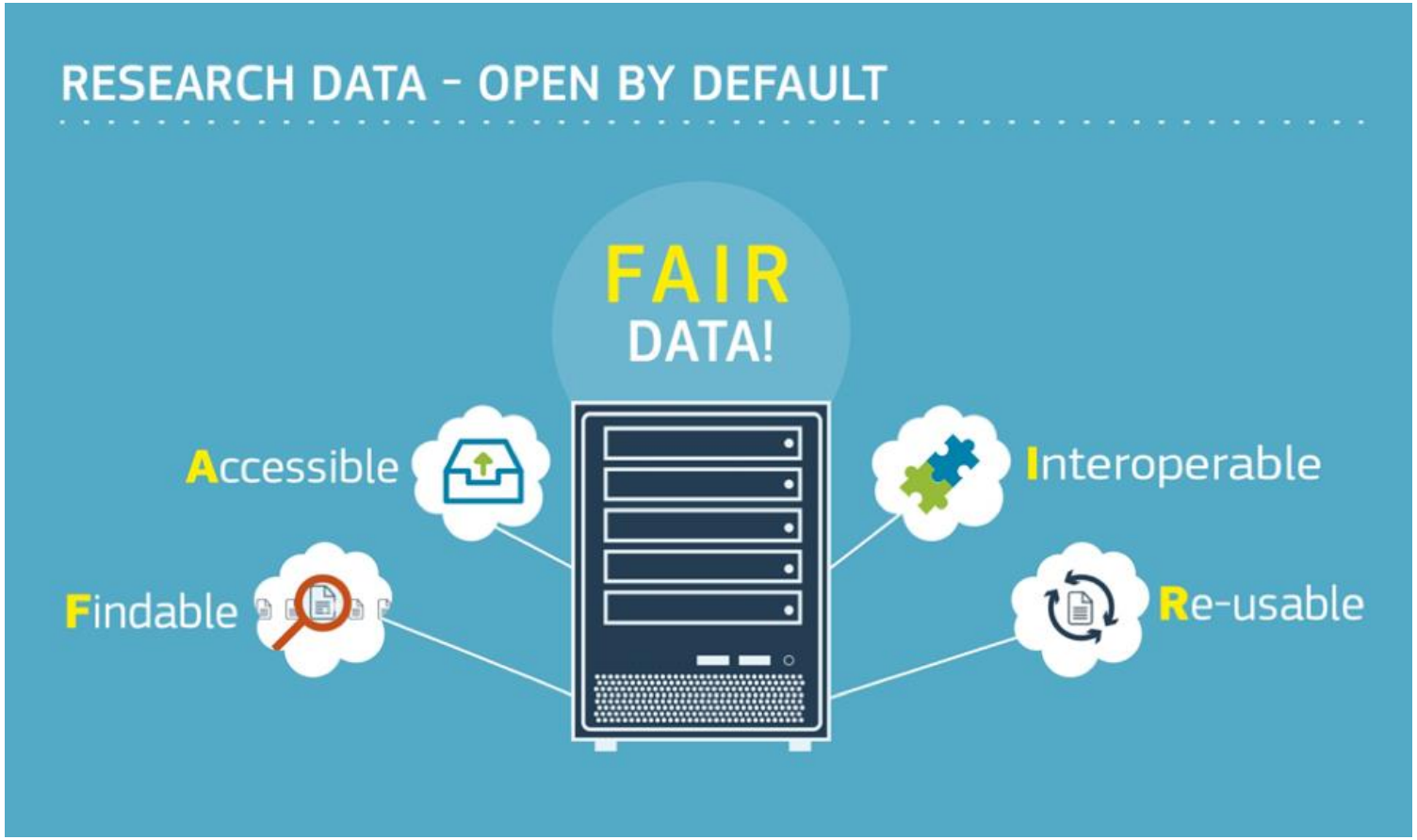
Group A: RDA outputs

Group B: RDA WG's and IG's related to sensitive data

Groups switch themes after 30 min

14.30-15.00 Workshop wrap-up

FAIR guiding principles from the point of view of sensitive data



Picture source: OpenAire https://www.openaire.eu/how-to-make-your-data-fair

Why Care?

For one because they do…

LIBER

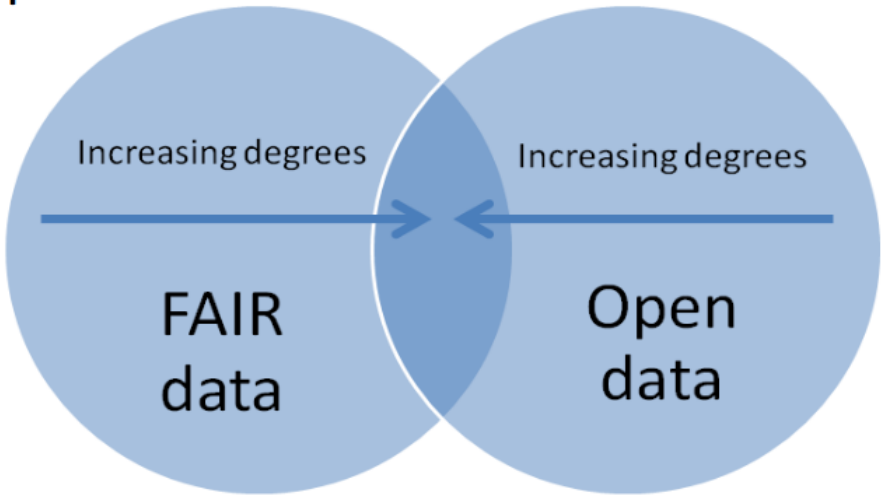
UNIFI

ALLEA
ALL European
Academies

G20 2016 CHINA

FAIR and Open

- Concepts of FAIR and Open should not be conflated.
- Data can be FAIR or Open, both or neither
- The greatest potential reuse comes when data are both FAIR and Open



Slide source: Sarah Jones, DCC https://library.ust.hk/wp-content/uploads/2019/03/FAIR-data.pdf

F and A

Findable

F1. Metadata are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. Metadata are registered or indexed in a searchable resource

Accessible

A1. Metadata are retrievable by their identifier using a standardised communications protocol

- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

I and R

Interoperable

I1. Metadata use a formal, accessible, shared, and broadly applicable
language for knowledge representation.

I2. Metadata use vocabularies that follow FAIR principles

I3. Metadata include qualified references to other metadata

Reusable

R1. Metadata are richly described with a plurality of accurate and relevant attributes

- R1.1. Metadata are released with a clear and accessible data usage license
- R1.2. Metadata are associated with detailed provenance
- R1.3. Metadata meet domain-relevant community standards

"Most types of reusable data that are expensive to produce now have purpose-built databases. FAIR principles dictate the publication of rich metadata to describe these data and to enable discovery of what is contained therein, even in the case of sensitive data that identify persons. The data fields and metadata schema should be accessible, together with the details of any access restrictions, whether or not the underlying data can actually be accessed. "

Nature Genetics editorial, VOLUME 48 | NUMBER 4 |APRIL 2016, https://doi.org/10.1038/ng.3544

Let’s discuss!

Social / legal / human interoperability

The different stakeholders aren't just not speaking the same language - there isn't even a common view as to what data actually is.

Full understanding of consent to publish (including how data will be published) for researchers and research participants during data collection.

Recognition.

Data sharing and opening/publishing of pseudoanonymised data.

Balance between usability and security

Investing into infrastructure in which researchers can safely and appropriately store and manage their research data (ideally so that this would make their life easier, not more difficult, as has been common with many archiving practices). At the moment the whole collecting-archiving-reusing -pipeline in Finland is entirely broken, even more so with sensitive data, and it is very unclear whether there is sufficient conversation across different involved parties, including universities, archives, scientific associations and infrastructure serving organisations. These are not issues with which individual institutions or individuals can come up with their solutions, but we must work together to enable our common work. We also have to acknowledge that whether data is sensitive or not is often only understood as part of the research process, so the solutions where the data is initially sorted into different categories and environments by this criteria will not hold very well.

Agreeing on process and conditions on how personal / health / sensitive data from different register holders / data controllers can be brought together on a joint secure processing platform for research projects.

repository and storage policies including rules of sharing. Lack of clarity what is sufficient level of anonymization and security from GDPR and national legislation points-of-view. Also, how to draft a sensitive data access application which doesn't limit new research questions but is still as accurately defined as possible?

Infrastructure / service ecosystem

Storing. Opening project metadata

Storaging and sharing of sensitive data.

National tools and services that encompass different scientific fields.

Where to find a secure environment to access AND analyze e.g. label the data securely.

Getting practical and secure solutions for transferring or sharing sensitive data with trusted parties. How to store and share sensitive data in user-friendly manner.

Lack of platforms

Proprietary Data

What to do with data from global media companies (eg. social media data)? GDPR says to store only those record that are needed, and only for a limited time period - what about longitudinal research?

The big issue is that sensitive personal data is posessed by big companies - they use it for boosting their business, and these data are not available for research or social good.

BIO -ENGINEER AT PHARMACY